Chapter XXV
Grid Enabled Surrogate Modeling

Dirk Gorissen
Gent University–IBBT, Belgium

Tom Dhaene
Gent University–IBBT, Belgium

Piet Demeester
Gent University–IBBT, Belgium

Jan Broeckhove
University of Antwerp, Belgium

ABSTRACT

The simulation and optimization of complex systems is a very time consuming and computationally intensive task. Therefore, global surrogate modeling methods are often used for the efficient exploration of the design space, as they reduce the number of simulations needed. However, constructing such surrogate models (or metamodels) is often done in a straightforward, sequential fashion. In contrast, we present a framework that can leverage the use of compute clusters and grids in order to decrease the model generation time by efficiently running simulations in parallel. We describe the integration between surrogate modeling and grid computing on three levels: resource level, scheduling level and service level. Our approach is illustrated with a simple example from aerodynamics.

INTRODUCTION

Computer based simulation has become an integral part of the engineering design process. Rather than building real world prototypes and performing experiments, application scientists can build a computational model and simulate the physical processes at a fraction of the original cost. However, despite the steady growth of computing power, the computational cost to perform
these complex, high-fidelity simulations are still enormous. A simulation may take many minutes, hours, days or even weeks (Gu, 2001; Lin et al., 2005; Qian et al., 2006). This is especially evident for routine tasks such as optimization, sensitivity analysis and design space exploration as noted below:

“...it is reported that it takes Ford Motor Company about 36-160 hrs to run one crash simulation. For a two-variable optimization problem, assuming on average 50 iterations are needed by optimization and assuming each iteration needs one crash simulation, the total computation time would be 75 days to 11 months, which is unacceptable in practice” (Wang and Shan, 2007, p1).

Consequently, scientists have turned towards upfront approximation methods to reduce simulation times. The basic approach is to construct a simplified approximation of the computationally expensive simulator, which is then used in place of the original code to facilitate Multi-Objective Design Optimization (MDO), design space exploration, reliability analysis, and so on (Simpson, 2004). Since the approximation model acts as surrogate for the original code, it is referred to as a surrogate model or metamodel.

While the time needed for one evaluation of the original simulator is typically in the order of minutes or hours, the surrogate function, due to its compact mathematical notation, can be evaluated in the order of milliseconds. However, in order to construct an accurate surrogate one still requires evaluations of the original objective function, thus cost remains an issue. The focus of this paper is to discuss one technique to reduce this cost even further using distributed computing. By intelligently running simulations in parallel, the “wall-clock” execution time, in order to come to an acceptable surrogate model can be considerably reduced.

We present a framework that integrates the automated building of surrogate models with the distributed evaluation of the simulator. This integration occurs on multiple levels: resource level, scheduling level and the service level. Each of these will be detailed below.

BACKGROUND

Surrogate Modeling

Surrogate models play a significant role in many disciplines (hydrology, automotive industry, robotics, ...) where they help bridge the gap between simulation and understanding. The principal reason driving their use is that the simulator is too time consuming to run for a large number of simulations. A second reason is when simulating large scale systems, for example: a full-wave simulation of an electronic circuit board. Electro-magnetic modeling of the whole board in one run is almost intractable. Instead the board is modeled as a collection of small, compact, accurate surrogates that represent different functional components (capacitors, resistors, etc.) on the board.

There are a huge number of different surrogate model types available, with applications in domains ranging from medicine, ecology, economics to aerodynamics. Depending on the domain, popular model types include Radial Basis Function (RBF) models, Rational Functions, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Kriging models (Wang and Shan, 2007).

An important aspect of surrogate modeling is sample selection. Since data is computationally expensive to obtain, it is impossible to use traditional, one-shot, full factorial or space filling designs. Data points must be selected iteratively, there where the information gain will be the greatest (Kleijnen, 2005). A sampling function is needed that minimizes the number of sample points selected in each iteration, yet maximizes the information gain of each sampling step. This process is called adaptive sampling, but is also
Related Content

Soft-Checkpointing Based Hybrid Synchronous Checkpointing Protocol for Mobile Distributed Systems
[www.igi-global.com/article/soft-checkpointing-based-hybrid-synchronous/52047?camid=4v1a](www.igi-global.com/article/soft-checkpointing-based-hybrid-synchronous/52047?camid=4v1a)

Load Balancing in Peer-to-Peer Systems
[www.igi-global.com/chapter/load-balancing-peer-peer-systems/36408?camid=4v1a](www.igi-global.com/chapter/load-balancing-peer-peer-systems/36408?camid=4v1a)

BAM-Chord: DHT-Based Bandwidth Adaptive Multicast System
[www.igi-global.com/article/bam-chord-dht-based-bandwidth/76923?camid=4v1a](www.igi-global.com/article/bam-chord-dht-based-bandwidth/76923?camid=4v1a)

Trust Calculation and Management in P2P and Grid Systems
[www.igi-global.com/chapter/trust-calculation-management-p2p-grid/40827?camid=4v1a](www.igi-global.com/chapter/trust-calculation-management-p2p-grid/40827?camid=4v1a)